Condition of Lake Erie Largemouth Bass Sampled in the ODOW Nearshore Community Survey 2013-2017

Alex Benecke

January 4, 2018

Data Prep

```r
Stock <- read.csv("Data/Clean-Data/largemouth-bass_Wr_Stock.csv") %>%
  filter(Year < 2017) %>%
  filterD(!is.na(Wr)) %>%
  arrange(Year,gcat)


Stock$fyr <- factor(Stock$fyr)
Stock$Year <- factor(Stock$Year)
```

```r
headtail(Stock)
```

```
##     fyr Year Site FID Weight        Ws         Wr Length lcat20      gcat
## 1    13 2013   10   9    807  874.1340   92.31995    387    380 preferred
## 2    13 2013   11   1    968  934.6786  103.56501    395    380 preferred
## 3    13 2013   11   8   1159 1196.8993   96.83354    426    420 preferred
## 408  16 2016   18   8    479  336.1388  142.50066    289    280     stock
## 409  16 2016   18  14    466  351.6072  132.53427    293    280     stock
## 410  16 2016   18  23    473  375.7265  125.88945    299    280     stock
##     Age SexCon Sex
## 1     4      8   2
## 2     3      3   1
## 3     3      3   1
## 408   2      3   1
## 409   2      3   1
## 410   2      8   2
```

```r
str(Stock)
```

```
## 'data.frame':    410 obs. of  13 variables:
##  $ fyr   : Factor w/ 4 levels "13","14","15",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Year  : Factor w/ 4 levels "2013","2014",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Site  : int  10 11 11 15 15 15 15 15 18 18 ...
##  $ FID   : int  9 1 8 167 NA 170 176 180 142 141 ...
##  $ Weight: int  807 968 1159 1144 927 982 1015 1000 942 941 ...
##  $ Ws    : num  874 935 1197 859 882 ...
##  $ Wr    : num  92.3 103.6 96.8 133.1 105.2 ...
##  $ Length: int  387 395 426 385 388 401 406 411 381 382 ...
##  $ lcat20: int  380 380 420 380 380 400 400 400 380 380 ...
##  $ gcat  : Factor w/ 3 levels "preferred","quality",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Age   : int  4 3 3 3 NA 4 5 5 3 3 ...
##  $ SexCon: int  8 3 3 3 3 8 8 8 8 3 ...
##  $ Sex   : int  2 1 1 1 1 2 2 2 2 1 ...
```

```r
unique(Stock$fyr)
```

```
## [1] 13 14 15 16
## Levels: 13 14 15 16
```

> **Note**
>
> *I removed the years 2012 and 2017. 2012 because only large fish have weight length data and more and differenct sites were samples. I removed 2017 due to differences in the survey.*

> **Note**
>
> *I am removing a fish from site 15 year 2013 because it appears to be a outlier (Wr > 200). Probably due to data entry error.* **I went back and did this in 'Create-Wr-Gabelhouse-Data.Rmd' where I make the data file I use for this analysis.**

Summarize Relative Weight by Year

```r
(Wr.Stock <- Summarize(Wr ~ Year, data = Stock) %>% arrange(Year))
```

```
##   Year   n     mean       sd   min     Q1 median    Q3   max
## 1 2013  97 113.3837 13.28398 76.14 104.40  113.0 122.2 143.8
## 2 2014 140 109.6148 15.74476 80.40  98.78  106.9 117.8 151.3
## 3 2015  67 109.8376 15.57996 78.22  98.53  108.6 120.8 149.8
## 4 2016 106 115.4942 13.83934 61.76 108.00  115.5 124.7 146.2
```

> **Note**
>
> *The relative weight data contains only stock length individuals. This is so that I can easily compare the relative weight of fish with PSD. This is done despite the min TL being 150 mm. I may want to summarize relative weight for 150mm and greater length individuals in the future to see if young/small fish drive down or increase Wr.*

Lets start exploring the relative weight data. I have two questions I would like to know the answer to.

1) does Wr differ among years?

2) does Wr differ among gabelhouse length categories?

First Lets see if Wr is different between years.

```r
aov1 <- lm(Wr ~ Year, data = Stock)
# save(aov1,file = 'model-output/aov1.rda')
Anova(aov1)
```

```
## Anova Table (Type II tests)
##
## Response: Wr
##           Sum Sq  Df F value    Pr(>F)
## Year        2588   3  4.0016  0.007933 **
## Residuals  87529 406
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(aov1)
```

```
##
## Call:
## lm(formula = Wr ~ Year, data = Stock)
```

```
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -53.733  -9.356  -1.165   9.300  41.716 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  113.384      1.491  76.054   <2e-16 ***
## Year2014      -3.769      1.940  -1.943   0.0527 .  
## Year2015      -3.546      2.332  -1.520   0.1292    
## Year2016       2.110      2.063   1.023   0.3069    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 14.68 on 406 degrees of freedom
## Multiple R-squared:  0.02872,    Adjusted R-squared:  0.02154 
## F-statistic: 4.002 on 3 and 406 DF,  p-value: 0.007933
```

```r
boxplot(Wr ~ Year, data = Stock)
```


3
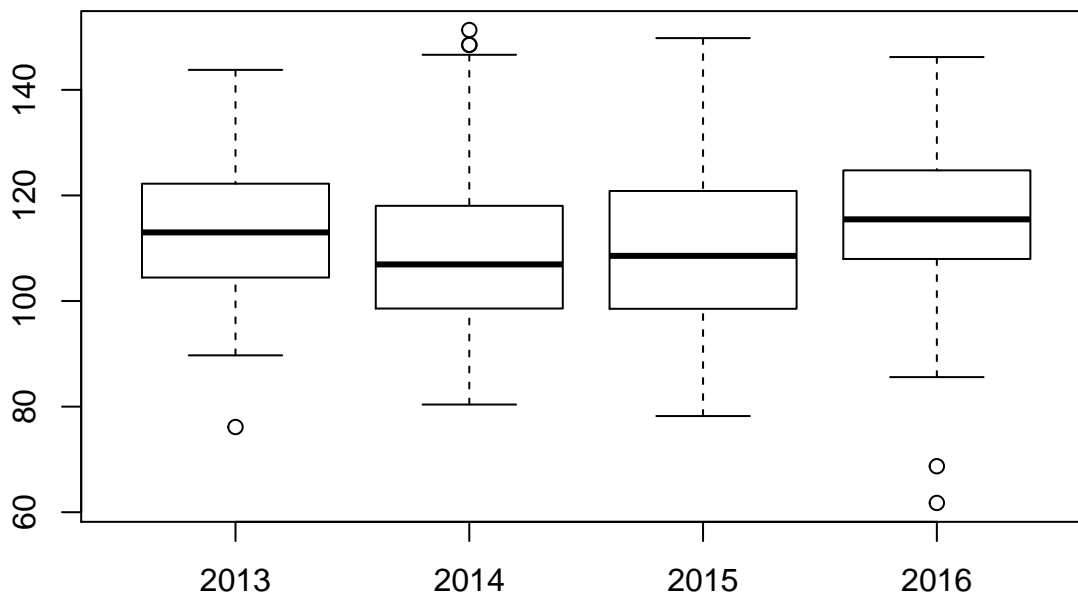
```r
mc1 <- glht(aov1, mcp(Year = "Tukey"))
summary(mc1)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = Wr ~ Year, data = Stock)
##
## Linear Hypotheses:
##                 Estimate Std. Error t value Pr(>|t|)
## 2014 - 2013 == 0  -3.7690     1.9397  -1.943   0.2103
## 2015 - 2013 == 0  -3.5461     2.3324  -1.520   0.4242
## 2016 - 2013 == 0   2.1104     2.0631   1.023   0.7346
## 2015 - 2014 == 0   0.2228     2.1812   0.102   0.9996
## 2016 - 2014 == 0   5.8794     1.8904   3.110   0.0108 *
## 2016 - 2015 == 0   5.6566     2.2916   2.468   0.0659 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Looks like Wr *is significantly different* between years (One-Way ANOVA, $F_{3,406} = 4.00$, p = 0.007933). There is *no significant difference* in relative weight between 2013 and 2014 (Tukey HSD, t = -1.94, p = 0.2101), 2013 and 2015 (Tukey HSD, t = -1.52, p = 0.4241), 2013 and 2016 (Tukey HSD, t = 1.02, p = 0.7346), 2015 and 2014 (Tukey HSD, t = 0.10, p = 0.9996), and 2015 and 2016 (Tukey HSD, t = 2.47, p = 0.0656). However, relative weight *is significantly different* between 2014 and 2016 (Tukey HSD, t = 3.11, p = 0.0107).

constructing a plot of Wr and Year

```r
grps.1 <- c("2013", "2014", "2015", "2016")
nd.1 <- data.frame(Year = factor(grps.1, levels = grps.1))
(pred.1 <- predict(aov1, nd.1, interval = "confidence"))
```

```
##        fit      lwr      upr
## 1 113.3837 110.4530 116.3144
## 2 109.6148 107.1753 112.0542
## 3 109.8376 106.3113 113.3639
## 4 115.4942 112.6906 118.2977
```

```r
plotCI(as.numeric(nd.1$Year), pred.1[, "fit"], li = pred.1[, "lwr"], ui = pred.1[,
    "upr"], pch = 19, xaxt = "n", xlim = c(0.8, 4.2), ylim = c(79, 121), xlab = "Year",
    ylab = "Mean Wr")

lines(nd.1$Year, pred.1[, "fit"], col = "gray50")

axis(1, at = nd.1$Year, labels = nd.1$Year)

cld(mc1)
```
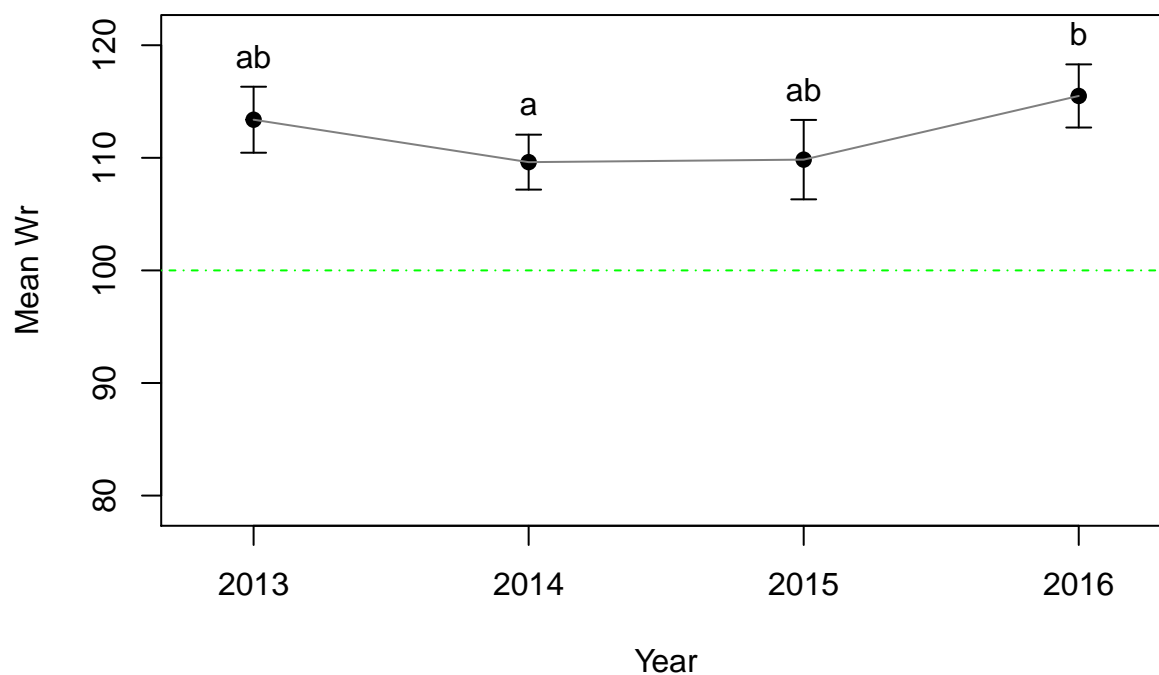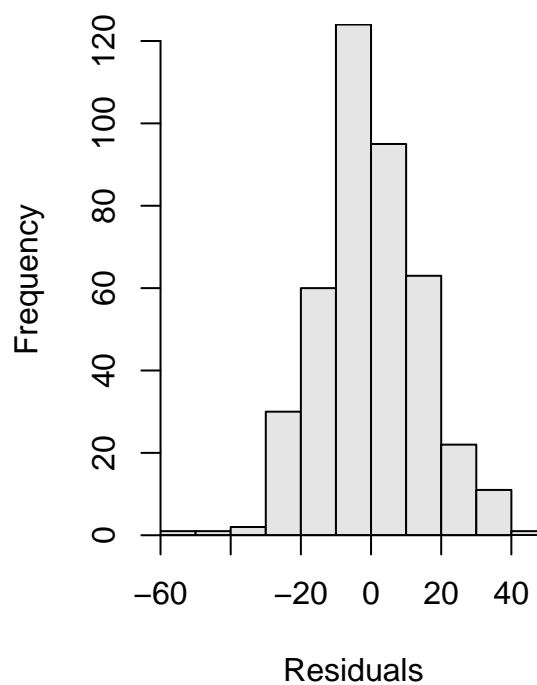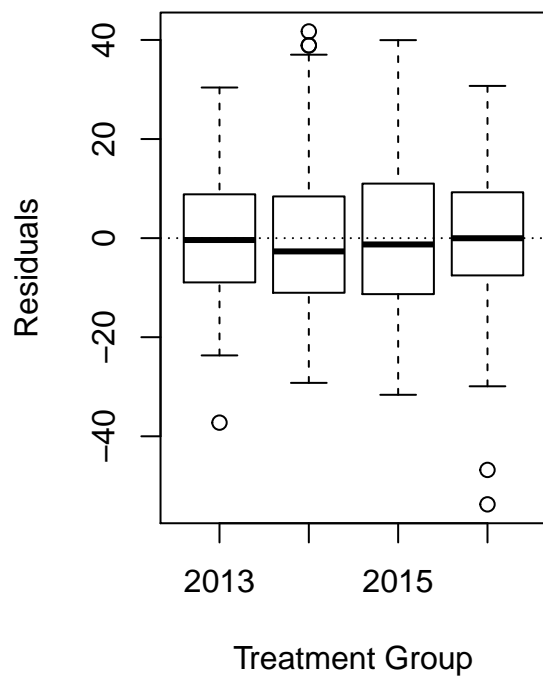
```
## 2013 2014 2015 2016
## "ab"  "a" "ab"  "b"
```

```r
text(x = nd.1$Year, y = pred.1[, "upr"], labels = c("ab", "a", "ab", "b"), pos = 3)

abline(h = 100, lty = 4, col = "green")
```



```r
residPlot(aov1)
```

```
leveneTest(aov1)    # Pitential outlier line 43
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   3  1.5423  0.203
##        406
# 2013 site 15 Weight = 1714 Length = 368 Wr > 200
```

Variance are equal and the homoscedasticity assumption is likley met (Levene's Test, $F_{3,406} = 1.5423$, p $= 0.203$).

Without 2017 *Variance are equal and the homoscedasticity assumption is likley met (Levene's Test, $F_{2,311} = 1.75$, p = 0.18).*

```
Year <- c("2013", "2014", "2015", "2016")
pred.1 <- data.frame(Year, pred.1)
names(pred.1) <- c("Year", "Wr", "LCI", "UCI")
str(pred.1)
```

```
## 'data.frame':    4 obs. of  4 variables:
##  $ Year: Factor w/ 4 levels "2013","2014",..: 1 2 3 4
##  $ Wr  : num   113 110 110 115
##  $ LCI : num   110 107 106 113
##  $ UCI : num   116 112 113 118
```

```
head(pred.1)
```

```
##   Year       Wr      LCI      UCI
## 1 2013 113.3837 110.4530 116.3144
## 2 2014 109.6148 107.1753 112.0542
## 3 2015 109.8376 106.3113 113.3639
## 4 2016 115.4942 112.6906 118.2977
```

```
# 2-9-2018#write.csv(pred.1,file =
# 'Data/Clean-Data/summary-data/relative-weight_largemouth-bass_STOCK.csv',row.names
# = FALSE)
```

2) Wr and Gcat

I will look into the difference in Wr between gcat at a later date. I don't think this matters so much as of now.

```
##  Year      gcat  n   mean    sd   min     Q1 median    Q3    max
##  2013 preferred 16 106.85 11.41  92.32  98.65 104.40 114.4 133.1
##  2014 preferred 18  97.67  8.94  83.65  90.40  98.58 103.0 115.5
##  2015 preferred 15  97.09 12.45  78.22  86.01  96.69 105.6 119.8
##  2016 preferred 10 107.72  7.17  94.36 104.30 107.80 111.7 118.9
##  2013   quality 40 108.20 11.40  76.14 103.60 107.70 116.6 131.3
##  2014   quality 57 103.51 11.64  80.40  96.07 102.20 111.6 133.1
##  2015   quality 38 109.32 12.92  84.02 100.80 108.10 120.5 131.0
##  2016   quality 44 111.22 14.36  61.76 105.50 110.50 118.9 146.2
##  2013     stock 41 120.99 12.03  94.13 113.20 120.80 128.4 143.8
```

```
## 2014     stock 65 118.27 15.78  88.74 106.70 116.10 127.5 151.3
## 2015     stock 14 124.90 12.56 103.80 116.60 123.90 133.9 149.8
## 2016     stock 52 120.60 12.54  68.71 113.90 121.60 127.1 144.9
```

IDK this doesn't look quite right below

```
gcat.aov <- aov(Wr ~ Year * gcat - 1, data = Stock)
# 3-7-2018#save(gcat.aov,file = 'model-output/gcat.aov.rda')

anova(gcat.aov)
```

```
## Analysis of Variance Table
##
## Response: Wr
##            Df  Sum Sq Mean Sq  F value Pr(>F)
## Year        4 5151406 1287852 7831.673 <2e-16 ***
## gcat        2   20569   10285   62.542 <2e-16 ***
## Year:gcat   6    1513     252    1.533 0.1659
## Residuals 398   65448     164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# summary(lm(Wr~ Year*gcat -1, data = Stock))
```

```
##                                 diff     lwr    upr p adj
## 2014:preferred-2013:preferred -9.181 -23.667  5.305 0.635
## 2015:preferred-2013:preferred -9.761 -24.913  5.391 0.610
## 2016:preferred-2013:preferred  0.871 -16.125 17.866 1.000
## 2015:preferred-2014:preferred -0.580 -15.319 14.159 1.000
## 2016:preferred-2014:preferred 10.051  -6.576 26.679 0.702
## 2016:preferred-2015:preferred 10.631  -6.580 27.843 0.672
## 2014:quality-2013:quality     -4.690 -13.386  4.006 0.832
## 2015:quality-2013:quality      1.117  -8.434 10.667 1.000
## 2016:quality-2013:quality      3.022  -6.188 12.232 0.995
## 2015:quality-2014:quality      5.807  -3.022 14.636 0.578
## 2016:quality-2014:quality      7.712  -0.748 16.173 0.113
## 2016:quality-2015:quality      1.905  -7.431 11.242 1.000
## 2014:stock-2013:stock         -2.714 -11.122  5.694 0.996
## 2015:stock-2013:stock          3.916  -9.135 16.966 0.998
## 2016:stock-2013:stock         -0.386  -9.191  8.419 1.000
## 2015:stock-2014:stock          6.629  -5.792 19.051 0.841
## 2016:stock-2014:stock          2.328  -5.516 10.172 0.998
## 2016:stock-2015:stock         -4.302 -16.996  8.392 0.994
```

There is *no significant difference* in mean Wr within gabelhouse length categories among years years ($F_{6,398} = 1.53$, p = 0.17).

```
par(mar=c(6.75,4,0.25,0.25))

boxplot(Wr ~ gcat + Year, data=Stock,
        at = c(1,2,3,5,6,7,9,10,11,13,14,15),
        xaxt="n", yaxt="n")

abline(h=100,lty=2,col="green")

mtext(c("Preferred", "Quality", "Stock","Preferred", "Quality", "Stock","Preferred", "Quality", "Stock"
        side = 1,
```
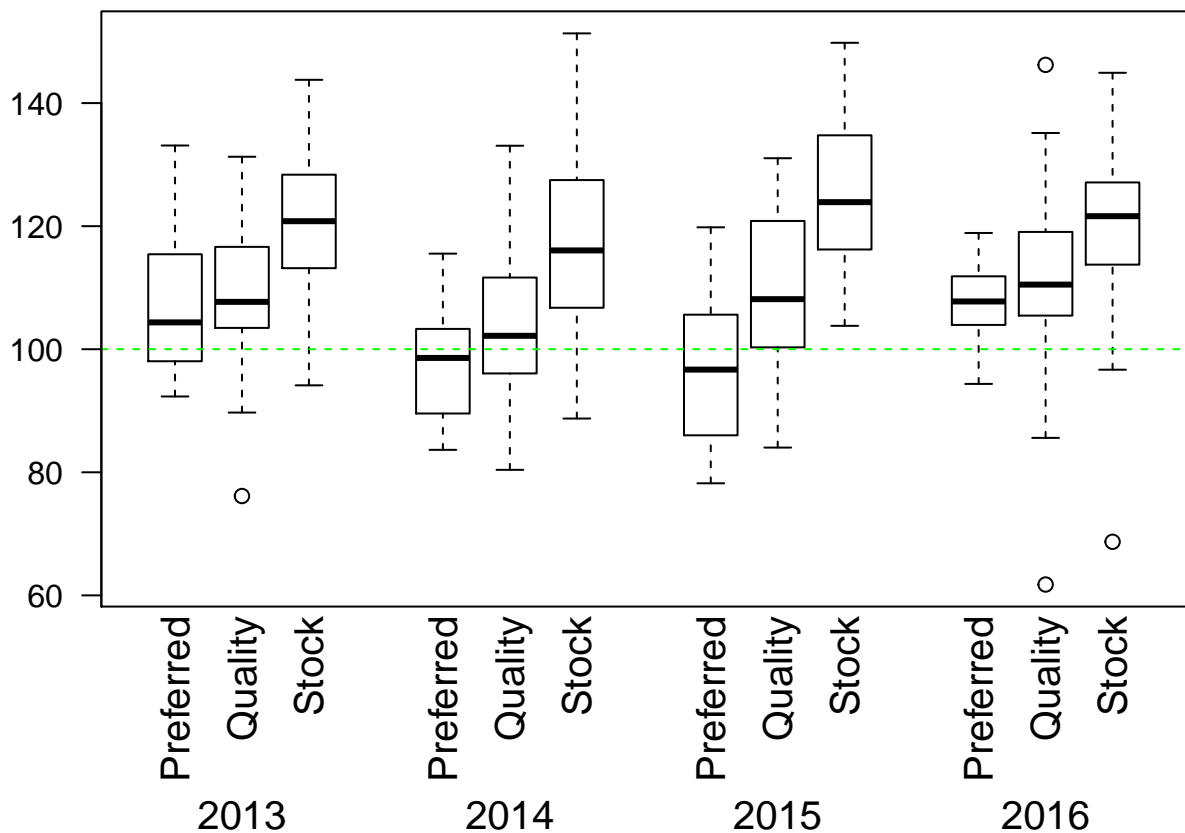
```
    line = 0.25,
    las=2,
    at = c(1,2,3,5,6,7,9,10,11,13,14,15),
    cex = 1.25)

mtext(c("2013", "2014", "2015", "2016"),
    side = 1,
    line = 5,
    at = c(2,6,10,14),
    cex = 1.25)

axis(2,
    at = seq(0,150,20),
    las = 2)
```


I want to see if Wr is different between gcat within each year. Above I show where the relationships of mean Wr for particular gcats between years.

```
Wr.13 <- filterD(Stock,Year==2013)
Wr.14 <- filterD(Stock,Year==2014)
Wr.15 <- filterD(Stock,Year==2015)
Wr.16 <- filterD(Stock,Year==2016)


str(Wr.13)

## 'data.frame':    97 obs. of  13 variables:
##  $ fyr  : Factor w/ 1 level "13": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Year : Factor w/ 1 level "2013": 1 1 1 1 1 1 1 1 1 1 ...
```

```
## $ Site  : int  10 11 11 15 15 15 15 15 18 18 ...
## $ FID   : int  9 1 8 167 NA 170 176 180 142 141 ...
## $ Weight: int  807 968 1159 1144 927 982 1015 1000 942 941 ...
## $ Ws    : num  874 935 1197 859 882 ...
## $ Wr    : num  92.3 103.6 96.8 133.1 105.2 ...
## $ Length: int  387 395 426 385 388 401 406 411 381 382 ...
## $ lcat20: int  380 380 420 380 380 400 400 400 380 380 ...
## $ gcat  : Factor w/ 3 levels "preferred","quality",..: 1 1 1 1 1 1 1 1 1 1 ...
## $ Age   : int  4 3 3 3 NA 4 5 5 3 3 ...
## $ SexCon: int  8 3 3 3 3 8 8 8 8 3 ...
## $ Sex   : int  2 1 1 1 1 2 2 2 2 1 ...
```

2013

```
Wr13.aov <- lm(Wr ~gcat, data = Wr.13)
anova(Wr13.aov)
```

```
## Analysis of Variance Table
##
## Response: Wr
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## gcat       2  4127.7 2063.87  15.141 1.994e-06 ***
## Residuals 94 12812.8  136.31
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Wr13.mc <- glht(Wr13.aov,mcp(gcat="Tukey"))
summary(Wr13.mc)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = Wr ~ gcat, data = Wr.13)
##
## Linear Hypotheses:
##                      Estimate Std. Error t value Pr(>|t|)
## quality - preferred == 0    1.351      3.454   0.391 0.918325
## stock - preferred == 0     14.137      3.441   4.108 0.000243 ***
## stock - quality == 0       12.786      2.595   4.928  < 1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

There is a *Significant* difference in mean Wr between gabelhouse length categories during **2013** ($F_{2,94} = 15.14$, $p < 0.001$). Mean Wr of **Stock** length largemouth bass is *significantly* different from **quality** (Tukey HSD, t = 4.93, $p < 0.001$) and **preferred** (Tukey HSD, t = 4.11, $p < 0.001$) length largemouth bass.

2014

```
Wr14.aov <- lm(Wr ~gcat, data = Wr.14)
anova(Wr14.aov)
```

```
## Analysis of Variance Table
##
## Response: Wr
```

```
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## gcat         2  9565.3  4782.7  26.322 2.121e-10 ***
## Residuals 137 24892.4    181.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Wr14.mc <- glht(Wr14.aov,mcp(gcat="Tukey"))
summary(Wr14.mc)
```

```
##
##    Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = Wr ~ gcat, data = Wr.14)
##
## Linear Hypotheses:
##                      Estimate Std. Error t value Pr(>|t|)
## quality - preferred == 0     5.841      3.644   1.603    0.243
## stock - preferred == 0      20.604      3.590   5.739   <1e-04 ***
## stock - quality == 0        14.763      2.446   6.035   <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

There is a *Significant* difference in mean Wr between gabelhouse length categories during **2014** ($F_{2,137}$ = 26.32, p < 0.001). Mean Wr of **Stock** length largemouth bass is *significantly* different from **quality** (Tukey HSD, t = 6.04, p < 0.001) and **preferred** (Tukey HSD, t = 5.74, p < 0.001) length largemouth bass.

2015

```
Wr15.aov <- lm(Wr ~gcat, data = Wr.15)
anova(Wr15.aov)
```

```
## Analysis of Variance Table
##
## Response: Wr
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## gcat         2  5625.3 2812.67   17.317 9.75e-07 ***
## Residuals   64 10395.2  162.42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Wr15.mc <- glht(Wr15.aov,mcp(gcat="Tukey"))
summary(Wr15.mc)
```

```
##
##    Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = Wr ~ gcat, data = Wr.15)
##
## Linear Hypotheses:
##                       Estimate Std. Error t value Pr(>|t|)
```

```
## quality - preferred == 0    12.228      3.886   3.146 0.006797 **
## stock - preferred == 0      27.813      4.736   5.873  < 1e-04 ***
## stock - quality == 0        15.585      3.984   3.911 0.000608 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

There is a *Significant* difference in mean Wr between gabelhouse length categories during **2015** ($F_{2,64} = 17.32$, $p < 0.001$). Mean Wr of **Stock** length largemouth bass is *significantly* different from **quality** (Tukey HSD, t = 3.91, $p < 0.001$) and **preferred** (Tukey HSD, t = 4.74, $p < 0.001$) length largemouth bass as well as **quality** and **preferred** (Tukey HSD, t = , p = 0.007) length largemouth bass.

2016

```
Wr16.aov <- lm(Wr ~gcat, data = Wr.16)
anova(Wr16.aov)
```

```
## Analysis of Variance Table
##
## Response: Wr
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## gcat        2  2763.1 1381.56  8.2031 0.0004946 ***
## Residuals 103 17347.2  168.42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Wr16.mc <- glht(Wr16.aov,mcp(gcat="Tukey"))
summary(Wr16.mc)
```

```
##
##    Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = Wr ~ gcat, data = Wr.16)
##
## Linear Hypotheses:
##                       Estimate Std. Error t value Pr(>|t|)
## quality - preferred == 0   3.502      4.546   0.770   0.7153
## stock - preferred == 0    12.880      4.481   2.874   0.0128 *
## stock - quality == 0       9.378      2.658   3.528   0.0017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

There is a *Significant* difference in mean Wr between gabelhouse length categories during **2016** ($F_{2,103} = 8.20$, $p < 0.001$). Mean Wr of **Stock** length largemouth bass is *significantly* different from **quality** (Tukey HSD, t = 3.53, p = 0.002) and **preferred** (Tukey HSD, t = 2.87, p = 0.013) length largemouth bass.